

AWARD NUMBER: W81XWH-12-1-0330

TITLE: Salivary Proteomic and microRNA Biomarkers Development for Lung Cancer Detection

PRINCIPAL INVESTIGATOR: David Wong, DMD, DMSc

CONTRACTING ORGANIZATION: University of California, Los Angeles
Los Angeles, CA 90095-2000

REPORT DATE: August 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE August 2014		2. REPORT TYPE Annual		3. DATES COVERED 1 Aug 2013 - 31 Jul 2014	
4. TITLE AND SUBTITLE Salivary Proteomic and microRNA Biomarkers Development for Lung Cancer Detection.				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-12-1-0330	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) David Wong, DMD, DMSc E-Mail: dtww@ucla.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California, Los Angeles Los Angeles, CA 90095-2000				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This is a lung cancer biomarker development project to test the hypothesis that there are discriminatory proteomic and miRNA biomarkers in saliva that can detect lung cancer with the aim to reduce the number unnecessary diagnostic workups (bronchoscopy) in patients with suspicious chest symptoms. Preliminary data is in place to support that our salivary biomarker technologies can discover and validate lung cancer biomarkers in saliva. The major goal is to perform a properly powered biomarker discovery and definitive validation of salivary proteomic and miRNA biomarkers for detection of lung cancer based on PRoBE design principles (prospective-specimen-collection and retrospective-blinded-evaluation). The outcome of this three-year proposal will be a panel of definitively validated non-invasive saliva-based proteomic and micro-RNA biomarkers for detection of lung cancer.					
15. SUBJECT TERMS Lung cancer, Early detection, Saliva, Biomarkers					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unclassified	18. NUMBER OF PAGES 9	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
1. Introduction.....	2
2. Keywords.....	3
3. Overall Project Summary.....	3-6
4. Key Research Accomplishments.....	6
5. Conclusion.....	6
6. Publications, Abstracts, and Presentations.....	6
7. Inventions, Patents and Licenses.....	6
8. Reportable Outcomes.....	6
9. Other Achievements.....	7
10. References.....	7
11. Appendices.....	7

Introduction

This is a lung cancer biomarker development project to test the hypothesis that there are discriminatory miRNA and proteomic biomarkers in saliva that can detect lung cancer with the aim to reduce the number unnecessary diagnostic workups (bronchoscopy) in patients with suspicious chest symptoms. Preliminary data is in place to support that our salivary biomarker technologies can discover and validate lung cancer biomarkers in saliva. The major goal is to perform a properly powered biomarker discovery and definitive validation of salivary proteomic and miRNA biomarkers for detection of lung cancer based on PProBE design principles (prospective-specimen-collection and retrospective-blinded-evaluation). The outcome of this three-year proposal will be a panel of definitively validated non-invasive saliva-based proteomic and micro-RNA biomarkers for detection of lung cancer.

Keywords: Lung cancer, Early detection, Saliva, Biomarkers

Overall project summary

This is the second year of this DoD CDMRP Lung Cancer Research Program (LCRP) Investigator-initiated Translational Research Award project titled "Salivary Proteomic and microRNA Biomarkers Development for Lung Cancer Detection".

The first year of this lung cancer biomarker development project was spent in the obtainment of regulatory (IRB) approvals from the two performance sites of the project, University of California Los Angeles and the Greater Los Angeles VA (GLA-VA), as well as with the Human Research Protection Office (HRPO) at the US Army Medical Research and Materiel Command (USAMRMC). These lengthy regulatory procedures unfortunately caused a year of setbacks delaying the initiation of our translational research study to develop salivary biomarkers for lung cancer detection. We have since obtained full approval of the informed consent changes and the HRPO of USAMRMC have approved the use of human subjects of this lung cancer biomarker development study. On November 15, 2013 we obtained approval from Dr. Sheilah Rowe, the Scientific Officer, that our project was delayed for one year and consider the need of an extension of the performance period.

This progress report contains the research accomplishment of the Specific Aims 1 and 2 as contained in the original Statement of Work.

Aim 1: Accrual of Lung Cancer and Control Subjects- Based on PProBE Design

Milestone 1: *Accrual of 1560 saliva samples from patients with suspicious chest symptoms. Based on current practice, we anticipate 624 lung cancers and 936 are cancer free patients at the Greater Los Angeles VA hospital (GLA-VA) procured based on the PProBE-study design.*

As of August 23, 2014 we have screened 2470 patients with chest symptoms at the GLA-VA (159% of the targeted enrollment of 1560). Of these 211 subjects were endoscoped and 92 were confirmed with diagnosed of lung cancer. Our original study design anticipated 624 lung cancer cases by the end of year 01 with nodular sizes on CT > 1cm. The lung cancer yield turned out to be 32 cases with nodular size >1 cm. This is much lower than anticipated and necessitated us to modify the study design for the biomarker discovery Aim.

We proposed to use 30 lung cancer and 30 non-cancer saliva samples for the biomarker discovery. In addition to the cancer status of these lesions, we have also correlated the tumor size of these lesions based on their CT data. This inclusion is of clinical relevance and impact since the ability to develop salivary biomarkers that can predict cancer from non-cancer patients will be clinically impactful. By examining the plot of sample size against the proportion of genes exceeding the power threshold, we estimate that the sample size of 30 per group (cases and control) will prove statistical power of at least 99% for 98% of the genes whose true effects exceed a fold-change of 2. Further increasing the sample size brings little improvement in power. Saliva from 30 lung cancer patients and 30 matched controls were used for the discovery studies. Controls were matched for gender, age, smoking history and ethnicity. This matching will ensure a distributional match on potential confounders, however we will not use a matched pair analysis plan.

Aim 2: Salivary miRNA and Proteomic Biomarkers Discovery, Statistical and Systems Approaches to Candidate Biomarkers Selection

Milestone 2: *Optimized salivary biomarker discovery technologies and a systems approach will be used to identify candidate miRNA and proteomic salivary biomarkers for lung cancer detection in a discovery cohort of 30 cases and 30 controls randomly*

selected from Aim 1. Salivary biomarker optimized data mining approach will be used to identify to candidate markers.

During this year we have made significant efforts to integrate the emerging technology of RNA-Seq to discovery miRNA in saliva of lung cancer patients. This technology will allow us to obtain un-parallel detailed information of known and novel miRNA in saliva that can be developed for lung cancer non-invasive biomarkers.

RNA-Seq of extracellular RNA (exRNA) in saliva

RNA-Seq is emerging technology to obtain the most detailed information of RNA in a biological sample. While we originally proposed to use the Taqman MicroRNA Array Card for saliva miRNA discovery, the significant advantages to use RNA-Seq for saliva miRNA discovery for known and novel miRNA is compelling. We published the first RNA-Seq study on salivary RNA using the SOLID™ system [Spielmann, 2012 #7452]. In this project, we will use Illumina sequencing systems. We have generated data that support the quality, reproducibility and feasibility of our approach. We have compared multiple library generation methods, constructed different types of libraries to capture the whole spectrum of exRNA in saliva, evaluated the reproducibility of our methods, and obtained a preliminary landscape of relative and absolute concentration of various types of exRNAs in saliva [Bahn, 2014 #8162].

Library generation. A number of RNA-Seq library construction methods have been developed in the literature. We have evaluated the performance of alternative methods, we used multiple commercially available kits (NEB, Illumina, Clontech and NuGen) targeting different types of RNA. A typical bioanalyzer profile of saliva exRNA is shown in Fig. 1. For each library, 500ng of total RNA was used as input. Importantly, predefined amount of synthetic spike-in RNAs were added into each RNA sample equivalently, which will serve as internal standards to evaluate library efficiency, reproducibility, to normalize data across different samples, and to calculate absolute RNA abundance. The synthetic RNAs were purchased from Exiqon and Life Technologies for small RNA-Seq and regular RNA-Seq, respectively. The synthetic RNA pool consists of many distinct RNA species (>40 for small RNA) to ensure abundance and sequence diversity. Since it is known that RNA from saliva is partially degraded with size between 20 and 200nt, we modified the library generation methods to exclude polyA selection and include a size-selection step favoring RNAs below 200nt. Depletion of ribosomal RNA was not carried out since it is known that saliva has relatively less rRNA compared to cellular RNA. Note that although the regular RNA-Seq spike-in RNAs were polyadenylated, the random priming method used in regular RNA-Seq still allows their usage as reference standards. Using these optimized steps for salivary exRNA, we have constructed to date 30 of the 60 saliva samples (30 lung cancer, 30 controls). All samples were randomized to minimize batched effect for the next step, RNA-Seq.

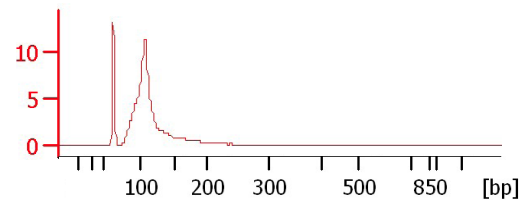


Figure 1. A typical bioanalyzer trace of total RNA from saliva isolated by Trizol LS.

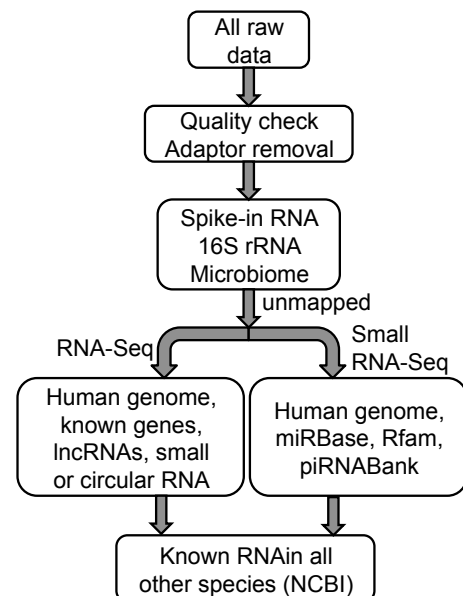


Figure 2. RNA-Seq and small RNA-Seq data analysis. Read mapping was carried out using Bowtie2 allowing up to 1 mismatch in the adaptor-trimmed reads.

Sequencing and Bioinformatic analysis of RNA-Seq data. All libraries were sequenced using Illumina HiSeq 2500 sequencers at the UCLA core facility. A total of 30-50 million single-end (50nt) reads were obtained for each library. We have developed customized bioinformatic pipelines to identify different types of non-coding RNAs (ncRNA) present in these data sets originated from human, microbiome, plants or any other species (Fig. 2). Small RNA-Seq data were analyzed for miRNAs and other ncRNAs. Although the small RNA libraries capitalize on the fact that canonical miRNAs have a 5'-phosphate and 3'-OH, other ncRNAs may be identified if their processing steps also lead to such footprints. Since the RNA-Seq libraries used random priming to generate cDNAs, they can theoretically capture all different types of RNAs in the selected size range (20-200nt) with adequate abundance. We have analyzed whether the RNA-Seq data sets may capture all long, small and circular ncRNAs. Mapping uniqueness was required for reads mapped to spike-in RNAs, known genes, lncRNAs and circular RNAs, but not for reads mapped to microbiome or 16S. Small RNA reads were not required to be unique either since small RNAs (miRNAs, piRNAs, etc.) may have multiple copies or similar family members in the human genome. All libraries yielded high quality reads, with an average of ~50% reads mapped to 16S and microbiome. To evaluate potential contamination by cellular RNA in our samples, we examined a number of genes (e.g., ESRP1/2, OVOL1/2, HBA1, APOC1 etc.) that are known to be highly specific to epithelial cells or leukocytes, the major types of cells in saliva. Most of these genes are not expressed based on the RNA-Seq data, supporting the effectiveness of our saliva SOP in removing cells.

Human miRNAs are abundant and stable in saliva. We have examined the reproducibility of miRNA profiling using the two different small RNA library generation methods.

An example of spike-in RNA correlation across libraries is shown in Fig. 3. Correlation coefficients (r) for all pairs of samples are mostly in the 0.97-0.99 range. Remarkably, all spike-in RNA species were found with reads, the abundance of which is highly correlated with their known relative abundance. Thus, we do not expect significant loss of RNA diversity during the library generation process. With this confirmed performance, we normalized across data sets using the spike-in RNA standards. A total of 332-418 miRNAs were identified in each library with at least 1 RPM. As expected, biological replicates were highly correlated in miRNA abundance (Fig. 3b). Remarkably, >300 miRNAs were common to the two donors (≥ 1 RPM in both) whose abundances are also highly correlated (Fig. 3c). Using the spike-in RNA, we estimated that the most abundant miRNA (miR-223-3p) has a concentration of 4.4pg per 1mL CFS. By sub-sampling the 30-50M reads for each library, we found that ~10M raw reads are adequate to enable robust miRNA (and other small RNA) quantification in general. Our data suggest that our methods are highly reproducible and a reference saliva miRNA profile is feasible. In addition, we observed that results from the two different library kits

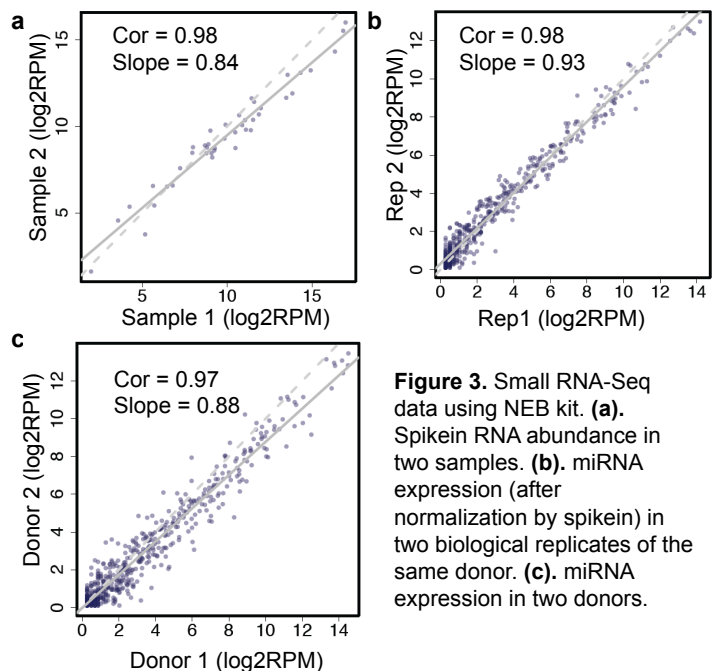


Figure 3. Small RNA-Seq data using NEB kit. **(a).** Spikein RNA abundance in two samples. **(b).** miRNA expression (after normalization by spikein) in two biological replicates of the same donor. **(c).** miRNA expression in two donors.

were very similar, with the NEB libraries yielding slightly more miRNAs and being relatively cost effective.

Key Research Accomplishments

- Accrual of 2470 patients with chest symptoms (159% of the targeted enrollment of 1560).
- Biomarker discovery cohort of 30 lung cancer patients with lung nodules on CT > 1cm and 30 non-lung cancer matched controls fully adhering to prospective-specimen-collection and retrospective-blinded-evaluation (PRoBE) design.
- Optimized RNA library construction and RNA-Seq technologies to saliva exRNA.
- Constructed RNA libraries to 30 saliva samples.

Conclusion

During the second year of the project, scientific progress has been sound. Targeted enrollment has been attained despite the lung cancer cases fulfilling the inclusion criteria was less than expected. Nonetheless we have proceeded with a biomarker discovery cohort of 30/30 that largely fulfilled the statistical power of the original proposal. We are particularly excited about the utilization of RNA-Seq for microRNA discovery in saliva for lung cancer detection. This emerging technology is most informative of the level and detail of salivary exRNA biomarkers to be harnessed. This will present a new frontier of salivary biomarker development that never existed before.

So What Section: The frontier technology of RNA-Seq for salivary miRNA development will lead to discovery of salivary biomarkers that will have discriminatory power to detect lung cancer in patients with symptomatic chest symptoms and nodules of > 1cm.

Publications, abstracts, and presentations

Bahn JH, Zhang Q, Li FL, Chan T-M, Lin XL, Kim YK, Wong DTW, Xiao XX. The Landscape of miRNA, piRNA, and Circular RNA in Human Saliva. 2014: pending revision at Clinical Chemistry.

Invention, Patents, and Licenses:

None to report.

Reportable Outcomes

Not applicable.

Other Achievements

None to report.

References

Not applicable.

Appendices

None.